- **RC1**:

General comments

The authors present the results of two climate model runs of different resolutions under idealized frameworks of aerosol forcing. The main focus of the analysis seems to be on the North Atlantic climate response to this external forcing at multidecadal timescales, and under low to medium resolution model configurations. The article is generally well-written and structured. The conclusions are supported by the analyses presented and I believe the research is novel and of substantial value to the climate research community. However, there are minor corrections to multiple sentences and to the figures that should be addressed, and further corrections to the explanation of processes that are mentioned.

Overall, the figures are generally too small, therefore, I included a separate section with advice on how to improve them. The authors should re-write the abstract to reflect the specificities of their analyses. It seems as some topics addressed in different sections are forgotten in the summary and conclusions, including explaining why the Labrador Sea is so important in their simulations, the multi-decadal variability analysis, and near-surface heat fluxes' imprint on net radiation and subsequent sea surface temperature.

Upon these corrections, I advise for publication.

Many thanks to the reviewer for carefully reading the manuscript and the thoughtful comments on improving the overall clarity of the text. I am very grateful for the positive review.

Specific comments

Title: Advise to include the word "multidecadal" after North Atlantic. - done

Abstract: From the abstract the main objectives seem to be to isolate the impact of Sulphur Dioxide onto North Atlantic SST variability and the effect of model resolution on the AMOC. However, the authors do not mention the multidecadal variability (particularly AMV) that is so present in their introduction. – The AMV is mentioned a few times in the introduction because that is the main way in which past studies have discussed multidecadal changes in SST across the whole North Atlantic basin. However, as this study and a previous study shows (Lai et al. 2022), the subpolar and subtropical North Atlantic SST are not necessarily linked and are driven by different processes. Hence, I have avoided mention of the AMV in the analysis and instead discussed SPNA SST by itself.

AMOC in density space vs. depth space: the analysis is centered already on the AMOC in density space, which is being recognized as beneficial to understand higher latitude processes. Perhaps the authors could leave the plots of the AMOC in depth space in the appendix to avoid confusion, and use the already detailed explanation as to why AMOC in density space was chosen. – Change wording because I have moved the AMOC in depth space to appendix

Multiple instances:

- replace "model diversity" with "model spread" - done

- replace "Labrador Sea ice" with "sea ice in the Labrador Sea" or "LS (as an acronym) sea ice" – revisit this again

- figures should be named with capita as Figure not figure or fig. Keep consistency. - done

- Consistency with SO2 being SO_2. - done

- The authors refer to density anomalies but do not indicate what density anomaly signal they refer to: positive density anomaly or negative. Therefore, when drawing conclusions on which processes contribute to specific buoyancy fluxes, one cannot know whether they contribute to denser or lighter waters, or in which regions.

- The authors use the word "change" in multiple instances without indicating the changes they refer to. It needs to be specified.

Lines 4-5: "large-scale cooling is followed by a slow…". What is the slow referring to here? the authors could add the simulation years to show the timescale, since it is not clear. – added lagging by 15-20 years

Lines 6-7: What happens to the density fluxes? Add an adjective to indicate how they are modified, and with respect to what – either the stronger AMOC or the higher resolution. – Reworded to 'This resolution difference is because sea ice growth in the Labrador Sea (LS) is stronger at low-resolution which inhibits air-sea interaction and reduces surface forced dense water formation, leading to a weaker AMOC response.'

Line 7: "The growth of Labrador Sea ice" – you refer to the sea ice in the Labrador Sea? If yes, replace with "sea-ice growth in the Labrador Sea is stronger". - done

Line 8: the surface buoyancy forcing here is always towards deep water formation, if yes, then it would lead to weaker AMOC. However, the buoyancy forcing can be either towards lighter or denser waters. Specify. – changed text to specify surface forced dense water formation

Line 17: Replace "decades of" with "at decadal timescales". - done

Line 26: What is your historical period? 1850-2014? – at this point I am not talking about modelling yet, so this refers to the historical period in the real world, i.e. up until now.

Lines 44-46: I do not see the link between "Furthermore, the modulation of the radiative budget is not the only way AA can drive SST changes;" and "turbulent heat flux changes can be driven by the advection of cold, dry air from the continents instead (Robson et al., 2022)". – changed text to emphasis that I am talking about local (modulation of radiative balance) vs remote driver (advection of aerosol cooled air from the continent)

Line 47: "Many of the beforementioned studies" and are these historical simulations based on CMIP6? Indicate what you mean by historical simulations. – It is both CMIP5 and CMIP6. The exact forcing is not important here, just that it is difficult to disentangle causal chains with these types of simulations.

Lines 49-51: This sentence needs to be clarified as it now only implies one modelling issue. - done

Line 54: "North Atlantic variability "change" in the coming decades". - done

Lines 56-57: How is using MM and LL relevant to this so-called second question? – reworded the second question to 'how model differences may affect the physical processes governing a model's forced response' to understand model spread

Lines 59-60: Replace model/models with version/versions. MM and LL are the versions of the same model. - done

Lines 62-66: As a personal taste, a paragraph with the article structure is not needed. – I am keeping it in because I find it useful sometimes to have a sense of the direction of analysis outlined from the outset

Line 69: "forms part" ?; and define CMIP. - done

Lines 75-79: Reformulate the paragraph. These are not model configurations but model components, and all of them are global, because that is your domain. Advice: "HadGEM3-GC3.1 is comprised of the GA/GL7.1 (Walters et al., 2019) atmosphere and GO6 (Storkey et al., 2018) ocean model components, coupled via OASIS-MCT coupler (Valcke et al., 2015), in timesteps 1 hour. Additionally, GA/GL7.1 implements the GLOMAP modal aerosol scheme, whereas the ocean component, based on version 3.6 of the Nucleus for European Modelling of the Ocean (NEMO) ocean model (Gurvan et al. 2017) also embeds the sea ice model GSI8.1 (Ridley et al., 2018)." – thanks for the suggestion, text has been changed as suggested

Line 81: Personal taste – remove "in CMIP6 nomenclature". - done

Line 82: This resolution is only in the mid-latitude atmosphere? – yes, specified in the text now

Line 90: Which parameter? – parameters used for model tuning

Line 92: Is ORCA025 eddy-resolving? Therefore, there is no GM parameterization? If yes, reorganize sentence to specify that GM is not applied in ORCA025 because of its resolution. – ORCA025 is not eddy-resolving, it is sometimes referred to as 'eddy-permitting'. At the latitudes of the Labrador Sea ORCA025 does not properly represent eddies, so GM is still applied.

Lines 108-109: This sentence should be somehow included in the abstract. - done

Line 114: "teleconnection mechanisms" – like? – 'such as Rossby wave train from the western North Pacific to Europe'

Line 128: Changes in? - 'linking AA to North Atlantic SST changes'

Line 131: Personal taste – remove "Lastly, for the PDRMIP simulations, the ocean outputs were not archived." – this information is not relevant to this study. - done

Line 132: It seems like this subsection of the methodology is part of the results section. – I decided to leave this in the method section so the result section focuses on the ocean response, to emphasis that this study is not about the aerosol and cloud processes.

Line 141: "indicates upward anomalies in Figure 2" - done

Line 143: Move "(3-7 days, Rasch et al. (2000))" to after "sulphate aerosols". - done

Line 159: "and subsequent AMOC strength" instead of "and also an indicator for how strong the AMOC is likely to be". - done

Line 160: Careful with causality here. The sentence implies that AMOC is directly modulated by TOA imbalance due to aerosol forcing, which is not the case. Perhaps restructure sentence to indicate to which extent it occurs. – reworded to 'This interhemispheric energy imbalance was found to correlate well with the magnitude of AMOC strengthening in CMIP6 historical simulations (Robson et al. 2022)'

Line 165: "0.25Wm-2" – superscript. - done

Lines 195-196: "stronger or weaker" instead of "a speeding up or slowing down" done

Line 198: Reformulate "Net DWSW is also sensitive to surface albedo changes, such as from changes in the sea ice." to detail the causality. - done

Line 200: How did you find out that THF and net DWSW explains more the SHF? – The sum of THF and DWSW explains almost all of SHF, the next largest term is the longwave (DWLW) which is an order of magnitude smaller than THF and DWSW. Clarified this in the text.

Lines 202-214: What is the physical meaning of the colors in your figures? A negative anomaly in surface heat fluxes indicates more heat loss from the ocean to the atmosphere, or from the atmosphere to the ocean? What is the direction of these fluxes? – Added the following sentence for clarification. Positive anomalies (red) denotes anomalous downward heat flux into the ocean and negative anomalies (blue) denotes anomalous upward heat flux out of the ocean and into the atmosphere.

Line 204: What is your domain for subpolar North Atlantic? – 45N to 60N, added to the text. This latitude range is consistent with later usage of SPNA, shown by the box in Figure 8

Section 2.3: How does the previous section connect directly into this section? This is missing. – The first paragraph of 2.3 explains this? Surface fluxes themselves can't explain the delayed warming, so we look at the ocean heat transport convergence and associated AMOC changes.

Line 216: To me it does not seem as if different sections of the North Atlantic respond differently to anything, but that surface fluxes are heterogeneously influenced by aerosol forcing in the North Atlantic. – Added text to clarify that the different sections of the N. At. is responding differently to the aerosol radiative fluxes rather than the total surface fluxes. i.e. that the North Atlantic wide DWSW changes can not explain the SPNA surface warming.

Line 219: Which changes? You haven't analysed the ocean heat content but ocean heat fluxes at the surface. – Added text to clarify that we have not yet looked at OHC changes, but we are linking the AMOC and OHT convergence changes to SPNA SST changes in subsection 2.3.

Line 221: "even as SO_2 emissions abruptly decrease after simulation year 30". - done

Line 226: replace "stronger warm" with "warmer". Additionally, the warmer anomalies are concentrated in the subtropical to subpolar North Atlantic. It is important to define your domain of subpolar North Atlantic.   - done. SPNA is defined previously as 45N to 60N, following a comment above.

Line 228: To which sigma is your AMOC in density space referenced? – Added text 'with respect to a reference density of 0 dBar'

Line 246: In what sense does the medium resolution have a stronger AMOC response?- Added text 'stronger AMOC response of ~3Sv compared to the lower resolution'

Line 267: W/m2 – consistency. Before you used Wm-2, even. So, it should be Wm^{-2}. - done

Line 303: It would be interesting to see the rest of the years in the appendix.

Lines 304: this has already been explained in lines 296-298. – removed this line

Line 314: Have you analysed interior mixing? See Sidorenko et al., 2020, 2021 for the role of interior mixing in the AMOC. – no I have not looked at interior mixing, but I have now added a reference to the suggested paper to point to its potential role.

Line 318: extremely dense with respect to what? Also, from the plots it seems that the marginal seas connected to the Atlantic are the ones with more outcropping. Not the Arctic. – Thanks for pointing this out, I meant the Barents and Kara Seas, not the Arctic. Added text to clarify these marginal seas are more dense than the SPNA.

Lines 323-326: Perhaps this conclusion changes if you look at density fluxes at the isopycnal of stronger AMOC, instead of at the surface. – I think this conclusion still holds even when considering the isopycnals at which the AMOC is strong. In Figure 8, the fluxes within the bounds of the isopycnal outcropping (clue and black contours) are still largely explained by the T component. I suppose this is less obvious in MM.

Line 339: either AMOC-sigma or AMOC-. Consistency is important here. - done

Line 345: density fluxes towards where? They have directions. Same in line 350 and other instances addressing fluxes. - The directions of the surface density fluxes are defined in Equation 1, at the start of section 2.4.2, where positive means surface heat and freshwater fluxes combined are acting to increase surface density.

Line 358: "Sea ice growth inhibits". Additionally, it does not inhibit air sea interactions but heat loss to the atmosphere. From figure 3 it also seems that there is a pronounced cooling in LL where sea ice grows more, not the opposite. – Clarified that it is heat loss to the atmosphere. Sea ice can also affect the surface albedo which can cool the surface so maybe this is another factor here. This would be a competing influence to the mechanism involving the shrinking isopycnal outcropping region and reduced surface density flux.

Lines 370-371: Again, inhibits heat loss, not air-sea interactions. – Clarified

Line 380: By surface forcing you mean heat forcing? To which changes and anomalies you refer to? AMOC strengthening, weakening, deepening, shoaling? Lighter, denser waters? – Changed text to highlight specifically that in this instance, anomalous heat fluxes are not the only explanation of AMOC strengthening, transport of dense water from elsewhere also matters. 'Local surface heat forcing is not the only driver of AMOC strengthening. Transport of anomalously dense water driven by surface fluxes in remote regions or from ocean circulation feedback can also affect subpolar North Atlantic density and hence the AMOC'

Line 385: "… Arctic. Jungclaus …" instead of "… Arctic, Jungclaus …". - done

Lines 386-388: Personal taste. The last sentence can be less inquisitive and show that this is an outstanding question. – left unchanged

Lines 392-397: The description of thermal and haline terms can be combined for conciseness. Both are similar but with one held constant, so having them described separately is redundant. – I feel this is fine for clarity if not conciseness. Personally I often get confused when people

combine explanations using constructs like 'The thermal (haline) component is calculated with the salinity (temperature) held at climatological value'

Lines 397-399: Why do you calculate the relative importance this way? What do you mean by square? Square root, T^2? Whatever it is also needs to be in the title (i.e. T^2-S). – Clarified to mean I compare the magnitude of T and S to assess their relative importance on the total density anomaly i.e. |T|-|S|.

Lines 409-411: "In the western and central SPNA, which includes the LS and IS, both models show positive density anomalies, although with different relative contributions from temperature and salinity." - done

Line 418: "From 1500-3000 m". - done

Line 433: Ocean advection of what? – text changed to highlight anomalous ocean-advection of salt in particular

Line 444: Move "resolutions" to after the parenthesis. - done

Lines 447-466: This is conclusions, should be at the end. – done

Line 448: SPNA is in the Northern Hemisphere. How can the Northern Hemisphere be lagged to itself? – changed text to clarify that the SPNA lags the rest of the northern hemisphere by 15-20 years.

Line 449: "which is more pronounced" instead of "with a stronger AMOC response" - done

Line 451: Double parenthesis. - done

Lines 455-458: Which differences? What do you mean by mean state here? Also rephrase the analyses of sea ice growth. It is a little odd in the sentence. – reworded this bullet point to clarify the role mean LS sea ice and SST on the AMOC response.

Line 473: Interrogate? – changed to investigate

Lines 476-478: It is the same model but tuned for different resolutions – changed wording to 'same model at two different resolutions with minimal tuning differences'

Line 486: It is unlikely so. Usage of large ensembles have their advantages. And your model is not necessarily better, or your setup is necessarily better. It is different. Careful with overselling the results. – good point. Changed wording to highlight differences in interpretation.

Line 501: You add more insight on how high resolution can also imprint its own bias due to the stochastic nature of climatic processes. Perhaps look at Hirschi et al. 2020. - done

Figures

Most figures could be plotted as a subplot to decrease the space between each subplot and improve representation, as figures are too small in general. To that, colorbars and axes which are shared should be omitted. In terms of timeseries, if the years are not dates (i.e. 1850-1900), it is simulation years. Additionally, some figures are plotted focused only on the North Atlantic, while others also show the global domain. This should be consistent unless there is a specific global analysis, which is not the case for SSTs, for example, when the authors only describe changes in the North Atlantic but plot it for the entire globe. Another common feature is the repetition of information in figure titles that is not necessary as they are already described in the

caption. – The plots showing a global domain are meant to provide a global context and set the scene, i.e. fig 1, and fig 2, is meant to show that there is a hemispheric wide aerosol forced signal, and that there is a unique signal in the North Atlantic. The next figures then focus on the processes specific to the North Atlantic.

Because of the way the plotting functions are written, and the subplot panels are constructed, it is difficult to decrease the space between the subplots without a lot of manual image editing. Where I can, I have reduced the number of columns on the multi-panel figures, so the figures are enlarged when fitted to the page.

Below is the specific description of suggested changes for each figure.

FIGURE 1:

1. a and b share the same colorbar. Use only one and increase figure by putting both with less horizontal space between them.

2. c and d share the y axes. Omit the y-axis label on d.

FIGURE 2: - done

1. a to f are all representative of annual means from years 10-30. Remove this from the title.

2. a and d share colorbars. Therefore, the colorbar should be under d, with the x-axis label omitted in a. Same holds for b and e, and c and f.

3. figures g, h and i are also annual and reflect the same domain. Remove this from the title. – Just personal taste but I feel it's clearer to leave this in

FIGURE 3: Compare processes rather than resolutions. This way one can easily see the difference between MM SST and LL SST one beside the other and the colorbars can be bigger and shared. Put the years the left as y-axis label. Figure will become bigger and differences will be easier to see. – Sorry, this is difficult to change this because of how my plotting functions are written. I have moved some superfluous variables to the appendix (AMOC-depth, thermal-haline decomposition of density fluxes), and reduced the empty spaces of figure 4. Hopefully that helps make the figures more legible.

FIGURES 4-11: Same as suggested for figures 1-3.

FIGURE 8: Why show the haline component in the figure if it is not in the same order of magnitude? Clearly the salinity contribution towards denser or lighter waters is not relevant here. I advise to plot the density flux at the isopycnal of maximum AMOC strength, there the haline component might play a role, instead of at the surface. Otherwise, the plot can be only of the total density flux, since it is just influenced by the thermal component. – Moved the decomposition into the to improve clarity

FIGURE 9: Consistency with MM and LL instead of N126 and N96. This is related to resolution but you have not used this convention in the methodology. - done

FIGURE 10: No need for MM-LL in the title. – Just personal taste but I feel it's clearer to leave this in.

FIGURE 11-12: No need to include the depth interval. This can be in the figure description. – Just personal taste but I feel it's clearer to leave this in

FIGURE 13: define -ve, +ve and so on. This figure is also quite convoluted. Perhaps be more specific as to which anomalies you refer (cooling, warming, salinization, freshening, and so on). – Removed +ve and -ve from the schematic and changed wording to make it more descriptive about increased cooling, increased surface density etc.